ESD-TR-81-342

DESIGN AND DEVELOPMENT OF A SEISMIC DATA CENTER

Michael A. Chinnery and Alvin C. Gann

Applied Seismology Group, Lincoln Laboratory, M.I.T.,
Cambridge, Massachusetts   02142, U.S.A.

Ann U. Kerr

Defense Advanced Research Projects Agency, Arlington,
Virginia   22209, U.S.A.

        The U. S. has embarked on the design and development of a
sophisticated data center for the analysis and management of
seismic data.   There are two broad motivations for this project.
First, we need to develop a new generation system for the
management and retrieval of digital seismic data, and to provide
a modern data resource for the research community.   Second, we
need a facility that will be able to respond to any U. S.
obligations that might be incurred under future agreements on
international seismic data exchange and global seismic monitoring,
such as those currently under discussion by the Committee on
Disarmament.   This paper discusses the functions that the data
center will be required to perform, and its basic computer
architecture.

INTRODUCTION

        The need for a modern seismic data center has arisen from
two recent developments. First, as seismology enters the
transition from reliance on analog paper and film data to full
use of high dynamic range digital data, the existing methods
for the storage and distribution of data are inadequate and
must be replaced by more sophisticated systems.   Second, inter-
national discussions that may lead to agreements on interna-
tional seismic data exchange and global seismic monitoring
will require the development of systems that will accept both

parametric data reports and seismic waveform data, perhaps in close to real time, and rapidly prepare an event bulletin for general distribution.

In response to these needs, the U. S. in developing a state-of-the-art Seismic Data Center (SDC) that will fulfill the following objectives in support of the seismic research community.

1. To provide a state-of-the-art seismic data management facility for the storage and retrieval of digital seismic data.

2. To provide a test bed for the evaluation of a variety of automatic signal processing techniques for digital seismic data.

3. To develop the most efficient techniques for the rapid preparation of a comprehensive event bulletin, using both digital waveform data and station seismic arrival reports.

4. To provide a variety of services to the research community, including both local and remote access to all archived data and event bulletins.

The objectives in support of international exchange of seismic data are as follows:

1. To provide a facility that will meet all anticipated requirements for international exchange of seismic data.

2. To provide a test bed for evaluating the hardware, software and processing techniques that may be required under contemplated agreements on international exchange of seismic data.

It is not possible, at present, to formulate a detailed set of the functions that this data cer : will be required to perform. New digital stations are still being installed, and considerable research is needed before we can fully understand the best ways to use this new type of data. In the international arena, agreements on global seismic monitoring have yet to be concluded, and we cannot predict the detailed requirements that these agreements will place on the center.

In this paper, therefore, we describe in rather general terms the types of data that the SDC will receive, the types of products that it must output, and the storage, retrieval

and processing facilities that must be available. We also
briefly describe the basic computer architecture of the proto-
type SDC now being developed at Lincoln Laboratory.

INPUT DATA CHARACTERISTICS

The Seismic Data Center will be required to accept data
from a wide variety of sources (see Figure 1). Three main
categories of data are anticipated. The first consists of
data reports or level I data (1). These are measurements made
by station operators on observed seismic arrivals, and include
a variety of parameters such as arrival time, amplitude and
phase identification. The second consists of digital waveform
data from both single stations and arrays. Some of these
stations may be U. S. participants in some form of International
Exchange of Seismic Data (IESD); for these stations, one task
of the data center will be to extract data reports from the
waveforms, and distribute them to all IESD participants. The
third may consist of actual lists of events, with preliminary
event parameters. Such lists are frequently prepared by array
stations and certain existing earthquake information organiza-
tions.

These data types will arrive at the SDC in a variety of
different ways, and with highly variable delays (see Figure
1). One significant task of the SDC will be to ensure that
each of these sets of data is properly entered into the SDC
data base.

The rates at which these data enter the system will place
an important constraint on the sytem design. A modern three
component digital seismic station may generate data at a rate
in excess of 2 kbits/second. If a network of such stations
were to feed continuous data into the Seismic Data Center,
this data source would dominate all others by far, in terms of
both processing and storage requirements. We have selected a
system design input rate of 125 kbits/second, since this would
appear to provide ample capacity for the forseeable future.
We also require that the system be able to handle twice this
data rate during catch-up periods (e.g., following a hardware
failure, and that the system could be adapted to an average
input rate of 250 kbits/second by simply adding more hardware,
without invalidating the existing hardware.

DATA STORAGE AND RETRIEVAL

Requirements on the data base management system concern
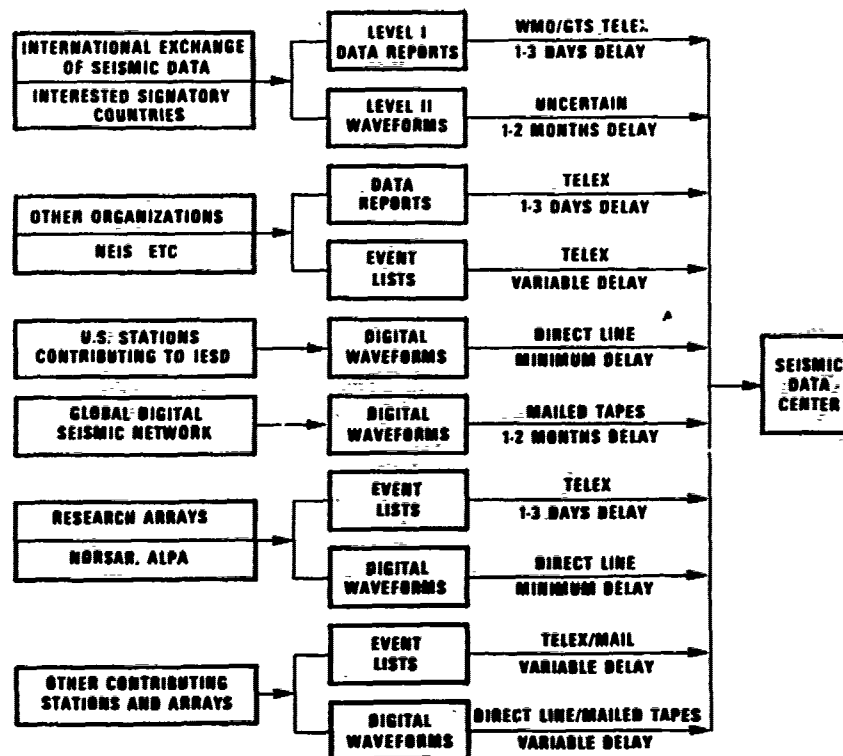its capacity and its speed of retrieval. The design data rate

Figure 1. Seismic Data Center Input Data Streams.

of 125 kbits/second corresponds to 10^{10} bits/day. This
requirement can be met by a combination of on-line disk
storage and off-line tape storage (commercially available
disk units can store about 5×10^9 bits, and high density
tape units can store about 10^9 bits on a 2400 foot tape). We
have selected this tape/disk solution for early SDC development.
We also recognize that other systems (such as optical disks)
may become available within the next few years. The SDC
architecture must have the flexibility that such devices can
be incorporated into the system with little problem at a
later stage.

The required rate of retrieval of data from on-line disk
storage depends somewhat on the type of processing carried
out at the SDC. We do not anticipate that there will normally
be a requirement to retrieve large quantities (say, all data
for an entire day) in one transmission. However, in order
not to exclude this possibility, we require that data can be

moved from disk storage to other parts of the SDC at a minimum
of an order of magnitude faster than real time (this corresponds
to a net signalling rate of at least 2 Mbits/second). With
this signalling rate, small quantities of data (say, a few
waveforms), can be retrieved from disk storage in a fraction
of a second.

Retrieval of data from off-line tape storage will inevitably
take longer. We, therefore, set the require.ent that, under
normal conditions, all input data will be stored on-line
until analysis of that data is completed. Retrieval from
tape will then be limited to a small number of requests,
primarily from research scientists, and we require simply
that all the waveforms for a given event be on a single tape.
The time needed to respond to a request for such data will
then be the time necessary to select a tape, mount it and
read it into the system. Depending on the total volume of
data input to the center, each tape will contain all available
data for a period ranging from 2 hours to 1 day.

In addition to routine functions, we also require that
the data base have the capacity to store certain special data
sets on-line. These would include reference events, used to
aid an analyst in seismic interpretation, and data sets
accumulated in response to a user request.

ROUTINE OUTPUT PRODUCTS

The SDC will produce a wide variety of routine output
data and provide a range of user services. These are summarized
in Figure 2.

There are two principal routine products. First, the
SDC must produce a comprehensive event bulletin. This will
provide a data base in itself for some users; to others, it
will function as an index to the waveform data stored at the
SDC. International discussions (1) have suggested that this
bulletin should be available with a net delay of 3-5 days
from real time. A bulletin produced this quickly will also
be of substantial use to the research community. However, we
note that some data may not be available at the SDC on this
time scale, and this suggests that it may be necessary to
prepare several bulletins. We require that the SDC be able
to compile an event bulletin within 24 hours of the receipt
of the last piece of input data used in its preparation (this
assumes 7 days a week SDC operation). This constraint will
satisfy all anticipated requirements.

The contents of the event bulletin are less well defined.

Certainly, the bulletin must contain all the event parameters
currently included in such publications as the Earthquake
Data Reports (of U. S. Geological Survey). We anticipate
that current and future research may lead to substantial
improvements in this list of parameters, and parameters
relating for example, to source mechanism may be included.
One immediate addition will be information about the availability
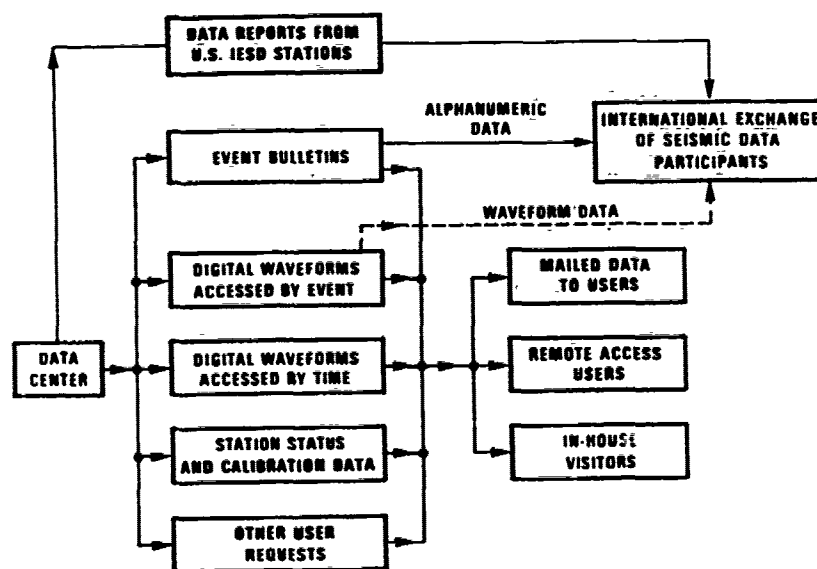of waveform data for listed seismic arrivals.

Figure 2. Data Center Products and User Interfaces.

The second important routine product will be access to
the waveform data base. We anticipate that user requests for
waveform data will be of two kinds. The user may request
waveform data for a particular event, as listed in the event
bulletin; or he may request the waveform data from a given
station (or stations) for a given time interval. Where
building up a large data set, he may leave a standing order
with the SDC for certain specific event types.

IESD output products will follow specifications laid out
in international agreements. At present, we anticipate
general distribution of data reports from U. S. stations
which are included in IESD and an event bulletin, both via
the WMO/GTS network. Level II waveform data will be supplied

by whatever methods are specified by the international agreements.

TIIE USER INTERFACE

     We plan to provide a variety of methods whereby the user
can interact with the data center. For those who can visit
the center, in-house computer facilities will be provided for
interaction with the data base, local interactive display and
analysis.

     For users who are unable to reach the center several
systems will be provided. . Data requests sent by mail or
telephone will be satisfied by mailed tapes or transmission
over a low data rate channel, such as a dialup phone connection.
The latter will be adequate for the transmission of event
bulletins, or small quantities of digital data.

     Where larger amounts of data are required, a more
sophisticated form of access would be needed for local record-
ing and computational facilities. We anticipate that the
capability required would range from a simple microcomputer
based terminal to something like a Seismic Analysis Station,
described later in this paper, depending on the level of
support required. Such a terminal would be connected to the
data center by a high data rate communications link.

     Certain types of analytical tools will also be made
available to the user. Examples include searching the event
bulletin on one or more of the event parameters, carrying out
filtering spectral analysis, rotation of horizontal components
and other kinds of time series analysis, and analytically
modifying the instrument response of the stored data (e.g.,
to produce a broad-band signal). Within the limitations of
system capacity, users will also be able to apply user-
generated processing methods to the data.

DATA PROCESSING AT THE SDC

     In order to translate the input data (Figure 1) into the
output data products (Figure 2), it will be necessary to
carry out a series of seismic processing steps within the
SDC. These processing steps will set some important require-
ments on the SDC design.

     Figure 3 shows, in schematic form, the processing steps
involved in the preparation of a final event bulletin. There
are two basic sets of procedures. The first is carried out
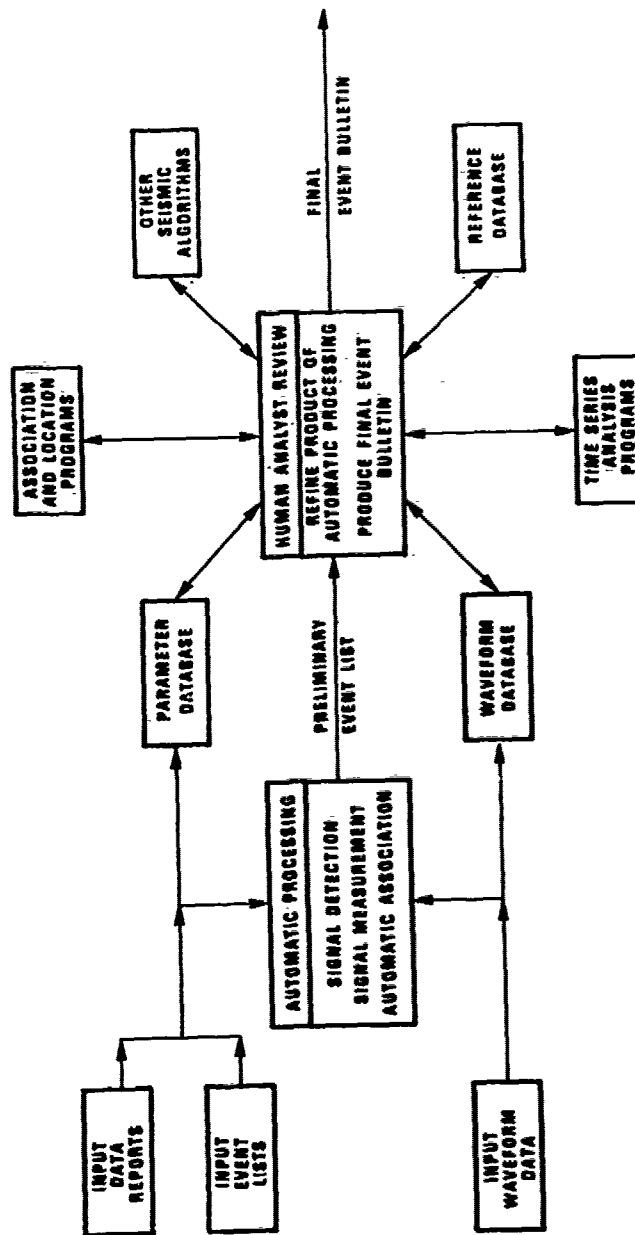completely automatically on the input data streams; the

Figure 3. Data Center Functions and Procedures.

second is carried out by a human analyst, who will refine the
results of the automatic processing steps using a variety of
software tools and with access to all available data.

The actual processing tasks that will be carried out
automatically are not yet clear. Many existing systems
include only signal detection in this category; however,
there are many indications in the literature that parameters
such as arrival time, amplitude, dominant period and signal
character can usefully be measured by the application of
automatic algorithms. A program of research is being carried
out in the U. S. with the objective of evaluating and implement-
ing these concepts. We anticipate that this will substantially
reduce the amount of human analyst time necessary in event
bulletin preparation.

Based on our present experience, however, there is no
way in which the human analyst can be eliminated from the
processing procedures. The subjective judgement and experience
of the mind are essential to the production of an acceptable
event bulletin. This places a requirement on the system for
a sophisticated interactive man-machine interface. The
analyst must be able to interact directly with event lists,
lists of signal characterization parameters, and digital
waveforms. He must be supplied with a wide variety of software
tools, including time series analysis programs, association
and location programs, and programs to aid him in the determina-
tion of event parameters such as magnitude and source mechanism.

Such an interactive "seismic analyst station" will
include all of the basic features that a researcher will need
in order to interact with the system. Our development of a
seismic analyst station is, therefore, aimed at devising a
very general purpose interactive seismic terminal, which
(with minor software variations) will satisfy the needs of
the operational analyst, the visiting research scientist and
the remote terminal operator.

SYSTEM ARCHITECTURE

The computer architecture which has been selected for
the impleme- .ation of the Seismic Data Center consists of
multiple minicomputers interconnected by a high data rate
local computer network. Such an arrangement provides great
flexibility, and easy expansibility to meet requirements that
may develop in the future.

Within this architecture, it is convenient to subdivide
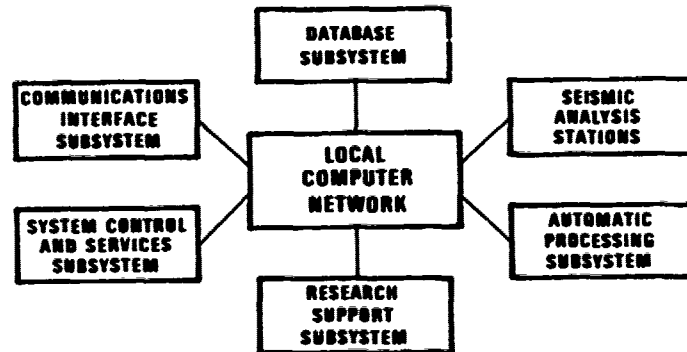the overall system into a series of subsystems, each of which

Figure 4.   Conceptual Seismic Data Center Configuration.

corresponds to a major functional component of the system
(see Figure 4). In particular, each subsystem will consist
of one or more computers with attached peripheral devices.
The principal features of these subsystems are summarized
below:

1. The Database Subsystem consists of two parts: the
    Parameter Database and the Waveform Database. The
    Parameter Database stores all of the alphanumeric
    data and provides a data management and retrieval
    system for events and arrivals. The Waveform
    Database provides a facility for maintaining and
    accessing the very large amounts of waveform data
    which the system acquires and processes.

2. The Research Support Subsystem will carry out a
    variety of computational tasks in support of research
    users, both local and remote. This support will
    include the selection of data sets for study, the
    analysis of both parametric and waveform data and
    the development of new algorithms and computer
    programs.

3. The System Control and Services Subsystem will
    carry out a variety of tasks, including the prepara-
    tion of data for distribution as requested by
    authorized outside users, development of new software
    for the system, and providing other basic system
    services.

4. The Automatic Processing Subsystem will carry out

automatic (non-interactive) computational tasks
related to detecting the arrival of seismic signals
from the digital waveform data and the automatic
association of arrival information into the prelimi-
nary characterization of events.

5. The Communications Interface Subsystem is responsible
    for all data input, and certain types of data
    output.

6. The Local Computer Network Subsystem interconnects
    the computers and provides all intercomputer communi-
    cation of data and control information. The local
    computer network is a key element in the system.
    It is required to be extremely reliable since it
    carries all intercomputer communications including
    all the data and all the command and status traffic
    in the system.

7. The Seismic Analysis Stations provide both seismic
    analysts and users with the ability to perform
    extensive computations on both the waveforms and
    the elements of the parameter database. Each of
    these is a single-user interactive computer system,
    with display devices, local input-output peripherals,
    and substantial disk storage.

Further details of these subsystems can be found elsewhere
(2).

CONCLUDING REMARKS

Initiatives arising from international discussions about
seismic data exchange and global seismic monitoring have
stimulated the design and development of a data storage and
retrieval facility that will fully utilize state-of-the-art
computer techniques. The challenge now is to ensure that
this data center will be fully responsive to the data needs
of the seismic research community. We seek the advice and
assistance of the research community, as we attempt to formulate
these requirements.

ACKNOWLEDGEMENTS

REFERENCES

1. *Report of the Conference of the Committee on Disarmament
    of the Ad Hoc Group of Scientific Experts to Consider
    International Co-operative Measures to Detect and to
    Identify Seismic Events,* Document CCD/558, Conference of
    the Committee on Disarmament, Geneva, 8 March 1978.

2. Chinnery, M. A., and Gann, A. G., *Design and Development
    of a Seismic Data Center,* Technical Note 1980-39,
    Lincoln Laboratory, M.I.T., 11 August 1980.